Privacy-Preserving and Explainable AI for Cardiovascular Imaging

Andrei PUIU[1,2], Anamaria VIZITIU[1,2], Cosmin NITA[1,2], Lucian ITU[1,2]*, Puneet SHARMA[3], Dorin COMANICIU[3]

[1]Advanta, Siemens SRL, 15th November Blvd, No. 78, Brasov, 500097, Romania
andrei.puiu@siemens.com, anamaria.vizitiu@siemens.com,
cosmin.nita@siemens.com, lucian.itu@siemens.com (**Corresponding author*)

[2]Automation and Information Technology, Transilvania University of Brasov,
5 Mihai Viteazu Street, Brasov, 500174, Romania

[3]Digital Technology & Innovation, Siemens Healthineers, 755 College Road, Princeton, 08540 NJ, USA
sharma.puneet@siemens-healthineers.com, dorin.comaniciu@siemens.com

Abstract: Medical imaging provides valuable input for managing cardiovascular disease (CVD), ranging from risk assessment to diagnosis, therapy planning and follow-up. Artificial intelligence (AI) based medical image analysis algorithms provide nowadays state-of-the-art results in CVD management, mainly due to the increase in computational power and data storage capacities. Various challenges remain to be addressed to speed-up the adoption of AI based solutions in routine CVD management. Although medical imaging and in general health data are abundant, the access and transfer of such data is difficult to realize due to ethical considerations. Hence, AI algorithms are often trained on relatively small datasets, thus limiting their robustness, and potentially leading to biased or skewed results for certain patient or pathology sub-groups. Furthermore, explainability and interpretability have become core requirements for AI algorithms, to ensure that the rationale behind output inference can be revealed. The paper focuses on recent developments related to these two challenges, discusses the clinical impact of proposed solutions, and provides conclusions for further research and development. It also presents examples related to the diagnosis of stable coronary artery disease, a whole-body circulation model for the assessment of structural heart disease, and to the diagnosis and treatment planning of aortic coarctation, a congenital heart disease.

Keywords: Artificial intelligence, Medical imaging, Cardiovascular disease, Explainability, Privacy preservation.

1. Introduction

Cardiovascular disease (CVD) is a major threat to human health and the leading cause of death worldwide (Thomas et al., 2018). Mortality and morbidity rates of CVD are increasing year by year, especially in developing regions. CVD generates a significant economic cost, estimated at 351.2 billion $ in the US, affecting the quality of life chronically (Virani et al., 2020). In the EU the yearly cost has been estimated at 210 billion €, divided between direct healthcare related costs (53%), productivity losses (26%), and the informal care of people with CVD (21%) (Timmis et al., 2018). An accurate and early evaluation of diagnosis and prognosis is crucial for improving and optimizing CVD outcomes.

Imaging plays a key role in every aspect of cardiovascular disease management, starting from baseline risk assessment diagnosis, staging, therapy planning, therapy delivery and follow-up. In addition, each type of heart disease (such as coronary artery disease, structural heart disease, arrhythmia, cardiomyopathy, congenital heart disease, cardiovascular toxicity, etc.) has led to the development of more advanced imaging methods and modalities to help the clinicians address the specific challenges in analysing the underlying disease mechanisms. These developments were driven by the need for a comprehensive quantification of cardiovascular structure and function across several cardiovascular imaging modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography, echocardiography, nuclear imaging. To address this need, researchers have been actively pursuing the development of advanced image analysis algorithms, some of which are routinely used in clinical practice (Mansi et al., 2019).

The majority of state-of-the-art image analysis algorithms are powered by artificial intelligence (AI) (Benjamins et al., 2019). The availability of unprecedented data storage capacity and computational power has allowed for the development, refinement and deployment of AI and specifically of machine learning (ML).

In CVD, AI algorithms have been successfully developed addressing various aspects ranging from the image acquisition level (e.g., scan workflow automation and efficiency) (Saltybaeva et al., 2018), to the reading and reporting level and prediction and prescription level (e.g., multi-scale modelling of the heart (Kayvanpour et al., 2015) and risk stratification in coronary artery

https://doi.org/10.24846/v30i2y202102

disease (Duguay et al., 2017)). For example, novel approaches have been introduced for the robust detection of anatomical structures, based on methods that reformulate the detection problem as a behaviour learning task for an artificial agent (Ghesu et al., 2019). Training of the artificial agent focused on distinguishing the anatomical object of interest from the rest of the body and on finding the object by learning and following an optimized navigation path towards the target object within the volumetric space. Recently also a novel method to achieve coronary artery labelling for structured reporting of Coronary CT Angiography (CCTA) was introduced (Fischer et al., 2020). The method relies on a deep learning (DL) model leveraging centerline labels annotated by experts for learning representations of coronary segments. Significant advancements have also been reported in the generation of patient-specific models of the mitral valve from medical images (Zhang et al., 2017). Several review articles have been published in recent years (Haq et al., 2020; Mathur et al., 2020), identifying challenges that need to be addressed to further increase the real-world adoption of AI based applications in the diagnosis and treatment planning of CVD. In the following the focus is on two main challenges.

ML relies extensively on existing and future patient data to deliver accurate and reliable results. Thus, the first challenge refers to the fact that, while biomedical data is abundant, it is hard to circulate and access due to ethical constraints, also affecting the development of computer-based solutions (Yan et al., 2019). There are concerns regarding protected health information related to patients (Mathur et al., 2020). Medical AI systems are difficult to realize, as data to develop and train them exist, but are locked inside hospital firewalls. To develop robust algorithms, the databases used for training, validation and evaluation should cover the entire spectrum of pathological variations and combinations. If training datasets lack diversity, algorithms may be biased or skewed to certain types of patients (Haq et al., 2020).

While this applies in general to all algorithms relying on medical data, a few specific aspects should be noted for CVD: significant geographic variations in CVD types and prevalence, and significant variations in imaging protocols, e.g., cardiac MRI (Fratz et al., 2013) and CT (Pulerwitz et al., 2020), have been reported.

Secondly, AI algorithms should be explainable and interpretable. ML algorithms are in general related to the concept of 'black-box', i.e., the rationale for how the outputs are inferred from the input data is unclear (Haq et al., 2020). Algorithmic decisions should however ideally provide a form of explainability (Bond et al., 2020). In general, explanations are about the attribution of the worth of input features towards the final model predictions, whereas interpretability refers to the deterministic propagation of information from input to response function. This challenge is consistent with recent clinically driven studies about transparency of clinical decision support (Richard et al., 2020) and agrees with a recent call for good practice in AI through models that are interpretable by design (Rudin, 2018). Similar to the first challenge, this applies to all algorithms employed in patient care, but, given the numerous types, subtypes and variations of CVD (Dey et al., 2019), the need for explainability and interpretability is even more pronounced in cardiovascular imaging.

This paper highlights the recent developments related to the above mentioned two challenges and discusses the potential impact of the existing solutions. Several examples are presented, related to the diagnosis of stable coronary artery disease, a whole-body circulation model for the assessment of structural heart disease, and to the diagnosis and treatment planning of a congenital heart disease. Section 2 addresses aspects related to privacy preservation in clinical AI applications, while explainability and interpretability requirements of an AI model are discussed in section 3. In the context of the approaches described herein, section 4 focuses on the impact of AI in clinical practice, and final conclusions are drawn in section 5.

2. Privacy Preservation in Clinical AI Applications

2.1 Privacy Concerns around Data Exploitation

Among all types of data asSoCiated with an individual, medical data has some of the highest privacy requirements, and the currently adopted regulations towards confidentiality guarantees for personal data manipulation (e.g., GDPR in EU, HIPAA in US) urges for the adoption of more effective privacy-preserving techniques (Shokri

& Shmatikov, 2015). Typically, to export sensitive data without compromising privacy, proper anonymization must be performed (Obermeyer & Emanuel, 2016). Thus, some of the data properties are modified, leading to a trade-off between privacy and utility. In the past few years, great effort has been invested in the development of different privacy preserving techniques with the potential of bridging the gap between data privacy and utility, demanded by the recent rise of privacy concerning scenarios. Cryptographic techniques such as Homomorphic Encryption (HE) offer a potential solution by allowing data to be encrypted while being manipulated (Kipnis & Hibshoosh, 2012). HE aims at keeping the data private by allowing a third party to process the data in the encrypted form without having to reveal the underlying information.

2.2 Homomorphic Encryption in Artificial Intelligence

Various researches have been conducted with the goal of performing privacy-preserving machine learning through homomorphic encryption. Orlandi et al. (2007) proposed the first notable approach for combining neural networks with HE. The method uses a cryptosystem that can only handle a few simple operations, the remaining operations being conducted through an interaction between the data owner and server. Neural networks fitted for inferring on encrypted data, such as CryptoNets (Dowlin et al., 2016), eliminate the interaction between the involved parties by using a Fully Homomorphic Encryption (FHE) scheme combined with polynomial approximation of non-linear functions. With focus on improving the neural network efficiency when operating on encrypted data, Chabanne et al. (2017) and CryptoDL (Hesamifard et al., 2017) further enhanced CryptoNet's capabilities by proposing different approximation strategies for the non-linearity property in neural network models.

The key downside of these privacy-preserving neural network solutions is the computational overhead: deeper networks need more computations, resulting in longer running time. In addition, an encoding strategy for floating point numbers has been used to empower computations on real-world data. Not only does the encoding strategy explicitly restrict the utility of these methods, but it also has a direct impact on the outcome of the computations. As a consequence, these methods use only encryption for the inference phase.

2.3 Privacy Preserving in Practice with Homomorphic Encryption

As a way of enabling computations to be performed on real data in practical medical applications, approaches that use some simpler homomorphic encryption cryptosystems based on linear transformations have been proposed in the specialized literature. This class of methods appears to be currently the only practical approach for performing privacy-preserving computations in real-world applications (Vizitiu et al., 2020). Consequently, Vizitiu et al. (2020) proposed a recent solution for privacy preserving deep learning based on a fully homomorphic encryption scheme. The standard steps are followed, i.e., the input data is encrypted and then sent to the server for training or prediction (Figure 1).

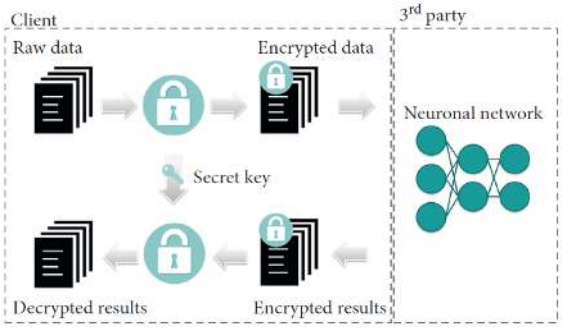

Figure 1. Workflow of the proposed privacy-preserving deep learning-based application relying on homomorphic encryption (Vizitiu et al., 2020)

The workflow relies on a variation of the Matrix Operation for Randomization or Encryption (MORE) encryption scheme, capable of operating directly on floating point data, allowing both training and inference similar to that of classical neural networks directly on homomorphically encrypted data. Evaluation was performed on the Modified National Institute of Standards and Technology (MNIST) dataset (a digit recognition benchmark problem) and on medical applications.

To demonstrate the feasibility of the proposed approach within CVD, a whole-body circulation (WBC) hemodynamic model of the cardiovascular system was chosen. The WBC model, displayed in Figure 2, contains in addition to the heart model, the systemic and pulmonary circulation. This hemodynamic model can determine different

clinically relevant quantities of interest under personalized conditions: arterial compliance, arterial resistance, ventricular/atrial/arterial elastance, dead volume of the left/right ventricle, pressure-volume loop, arterial ventricular coupling, etc. To ensure that model results are patient-specific the model parameters are calibrated iteratively based on the non-invasive measurements. This leads to an execution of up to one minute for computing the patient-specific quantities of interest. Thus, the authors investigated the possibility of training a deep learning model, under encryption conditions, for computing in real-time the measures of interest output by the WBC model.

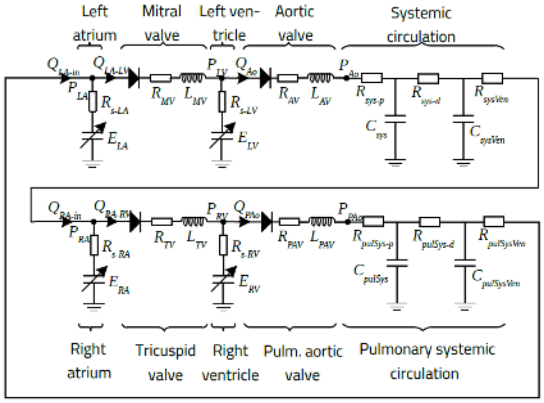

Figure 2. Whole-body hemodynamic model (Vizitiu et al., 2020)

To train such a model a database of synthetically generated patient-specific samples has been considered. For each sample, the personalization framework was run with the WBC model to assess the output measures of interest necessary to train the deep learning-based model (Vizitiu et al., 2020). A schematic of the workflow is displayed in Figure 3.

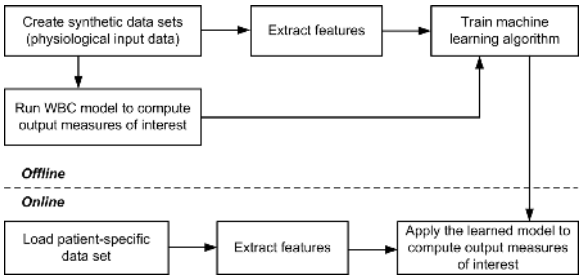

Figure 3. Proposed deep learning workflow relying on synthetic data (Vizitiu et al., 2020)

2.3.1 Methodology

In the MORE encryption scheme, a symmetric key is employed, and each numerical value is mapped to a matrix following the encryption. The MORE cryptographic approach is displayed in Figure 4 and the fully homomorphic behaviour is obtained through matrix algebra. Thus, the secret key encrypts the training data, and, next, the deep neural network is trained using only the ciphertext data, while the plaintext data is kept private at the data provider.

Message	Scalar value $m \in \mathbb{R}$
Secret key generation	Invertible matrix $S \subset \mathbb{R}^{2\times 2}$
Matrix construction	$M = \begin{pmatrix} m & 0 \\ 0 & r \end{pmatrix}$, where $r \in \mathbb{R}$ is a random parameter
Encryption operation	$Encryption(m) = C = SMS^{-1}$
Decryption operation	$Decryption(C) = K = (S^{-1}CS)$
Message recovery	$m = K_{(1,1)}$

Figure 4. MORE encryption scheme setup for rational numbers (Vizitiu et al., 2020)

A typical training pipeline is employed to train the model on the encrypted data, with support for floating-point computations, and with all mathematical operations being performed in the deep neural network on ciphertext data. Finally, a model is obtained which outputs ciphertext predictions, which can be interpreted only by the party which has the secret key. Importantly, during inference, the inputs need to be encrypted with the symmetric key that was also used to encrypt the training data.

2.3.2 Results

An identical accuracy was obtained for the medical and for the MNIST datasets by the plaintext and ciphertext deep learning models. Table 1 lists for the WBC model the MAPE (Mean Absolute Percentage Error) and the correlation (Pearson) coefficient obtained by evaluating the outcomes of the encrypted deep learning model, together with a statistical analysis demonstrating that the predictions are identical when being performed on encrypted and plaintext data. Following Levene's test for equality of variances, an F-statistic of zero was obtained, with a p-value close to 1. Thus, the test assumes equal variances between the predictions provided by the encrypted and plaintext WBC model. Therefore, the results of an independent samples t-test with the assumption of homogeneity of the variances were reported in Table 1. For all WBC parameters, the p-value provided by the t-test was close to 1, which implies that the means in every group were equal. In particular, these results indicate that there is no difference in performance between the encrypted and the plaintext WBC.

Table 1. Deep learning model results for the whole-body circulation model

Circulation	Parameters	MAPE (%)	Pearson corr. (%)	Levene's Test for Equality of Variances F-statistic × 10^{-16} (p-value)	t-test for equality of mean T-statistic × 10^{-16} (p-value)
Systemic	Dead volume [ml]	7.03	0.9997	0.2 (0.999)	0.1 (0.999)
	Time at max. elastance [s]	0.13	0.9995	0.2 (0.999)	0.4 (0.999)
	Resistance [g/(cm^4·s)]	0.17	0.9999	0.1 (0.999)	0.4 (0.999)
	Compliance [10^{-6} cm^4·s^2/g]	2.45	0.9867	0.4 (0.999)	0.2 (0.999)
	Ratio of prox. to distal resistance	1.36	0.9782	0.3 (0.999)	0.7 (0.999)
Pulmonary	Dead volume [ml]	9.88	0.9991	0.6 (0.999)	0.2 (0.999)
	Time at max. elastance [s]	0.10	0.9994	0.2 (0.999)	0.1 (0.999)
	Resistance [g/(cm^4·s)]	0.32	0.9998	0.5 (0.999)	0.3 (0.999)
	Compliance [10^{-6} cm^4·s^2/g]	0.67	0.9983	0.5 (0.999)	0.1 (0.999)
	Ratio of prox. to distal resistance	0.18	0.9999	0.4 (0.999)	0.6 (0.999)

The ciphertext solution leads to a larger runtime when compared to the plaintext version, but the overhead is significantly smaller compared to that of classic FHE schemes (Table 2).

Table 2. Execution times expressed as mean values ± standard deviations for the plaintext and ciphertext deep neural network models trained for performing the hemodynamic analysis

Operation	Runtime (s) on ciphertext data	Runtime (s) on plaintext data	Encrypted – Unencrypted ratio
Training (1 epoch)	0.66 ± 0.09	0.021 ± 0.001	31.4
Inference (2000 samples)	0.102 ± 0.01	0.006 ± 0.0009	17

The WBC model allows for a detailed assessment of time-varying measures of interest during one heart cycle. Pressure-volume (PV) loops, computed in real-time, represent one clinically relevant example. The left ventricular (LV) PV loop allows for a through characterization of the cardiac function. Different measures of the systemic circulation and the heart like cardiac output, stroke, volume, myocardial contractility, ejection fraction, cardiac oxygen consumption can be quantified therein. Other aspects like the ventricular-arterial mismatch, the degree of ventricular remodelling, and the LV end-diastolic PV relationship (Spevack et al., 2013) are correlated with congestive heart failure. Dilated cardiomyopathy, left ventricular hypertrophy and mitral and aortic valve regurgitation and stenosis (Hall, 2011) are pathologies which all induce changes in the PV loop.

2.4 Other Approaches

Although there are many promising recent studies for employing HE in AI workflows, it remains a difficult task, displaying significant limitations. To this extent, other approaches that are not relying on HE, have been also proposed. The original data may be secured using a special obfuscation operation that hides the sensitive components while preserving the statistics, such that it remains usable in a DL workflow. Obfuscation may be performed by adding a special type of noise (Romanelli et al. 2020) or by employing adversarial networks (Gong & Poellabauer, 2018; Abadi & Andersen, 2016).

2.5 Conclusions

CVDs are characterized by significant complexity and geographic variations in types and prevalence. Moreover, data acquisition in general, and medical imaging protocols in particular vary by region, vendor, etc. On the other hand, the development of robust AI algorithms for CVD requires access to large amounts of data from numerous sources. The approach described in this section and depicted in Figure 1 provides a practical solution to the challenges faced by large data collection initiatives in the context of strict privacy regulations.

The development of highly accurate AI algorithms requires significant effort related to problem understanding, data pre-processing and

filtering, model definition, results analysis, etc. These activities can only be performed on non-encrypted data. Hence, practically speaking, the following scenario might be imagined: the 3rd party in Figure 1 is first granted access to a non-encrypted dataset which is used for setting up data pre-processing pipelines, defining the AI model architecture, etc. to ensure an optimal prediction performance. In the second step (fully automated), the model is trained on large scale encrypted data collected from multiple sites, thus ensuring robust performance across all types of data.

3. Explainable Clinical AI Applications

User interfaces with explanations should accompany all AI algorithms, to display the rationale for interrogatable and transparent decisions taken by the algorithms, and to increase trust. Automation bias, i.e., naively trusted and accepted advice given by a machine (Bond et al., 2018), may also be mitigated by exposing such explanation user interfaces. Methods for explaining machine learning algorithms can be categorized into:

- Feature attribution: attributing the classification to a small number of numeric / semantic features. These algorithms are usually interpretable by design (Thomas et al., 2018; Naghavi et al., 2017);

- Saliency maps: sparse components of the original signal are identified, that have most influence on the model predictions, e.g., Local Interpretable Model-Agnostic Explanations – LIME (Virani et al., 2020);

- Activation maximization: for example, based on Generative Adversarial Networks (Romiti et al., 2020);

- Metric learning: it consists of deriving a metric from a classifier and using it to map out the data structure (LeCun et al., 2015). Additionally, explicit Siamese Networks have become very popular recently (Bertinetto et al., 2016).

In general, deep learning models are much more straightforward to apply as they eliminate the need for specific feature extraction methods, but they may be harder to interpret, and require much more training data as they rely on the optimization of a large number of parameters.

CVDs are characterized by numerous types, subtypes and minor variations (Dey et al., 2019), which are continuously updated in the clinical guidelines. Hence, to obtain trust, it is important for AI algorithms to not only capture these subtle differences but to present them to the clinicians.

In the following the uncertainty quantification is referred to as a concept closely related to explainability (subsection 3.1) and to the use of computational models in combination with machine learning models to enhance output explainability (subsections 3.2 and 3.3).

3.1 Uncertainty Quantification

Uncertainties are an inherent part of the world and they are inevitably propagated to data-driven systems. Uncertainty is also closely related to explainability and interpretability since AI model results displayed together with uncertainty measures potentially increase the trust, the acceptance and the adoption of the model in routine clinical practice.

Different types of uncertainties can be distinguished in AI applications:

- Data uncertainty: label noise (inherent class overlap) - a property of the underlying problem;

- Model uncertainty (epistemic uncertainty) - model complexity and out-of-sample behavior;

- Distributional uncertainty - systematic difference between the distribution of the training data and the distribution of the test data (Quionero-Candela et al., 2009).

Recently, a method for quantifying data and model uncertainties was proposed, yielding superior results compared to the state-of-the-art on three medical image classification tasks (Saltybaeva et al., 2018). Along with the class probability, the model estimates the uncertainty level reflecting the prediction confidence: a strong correlation between high predicted uncertainties and miss-labeled data has been demonstrated through a multi-radiologist-consensus user study. Uncertainty modelling can thus increase the trustworthiness of automated systems, making them more confident in their predictions while being able to identify uncertain situations requiring input from the clinician.

Distributional uncertainty can be detected using methods such as Normalizing Flows (NF)

(Kobyzev et al., 2020) in a pre-processing step. In the NF framework, an invertible mapping is learned between samples x ~ p(x) and latent variables z~ p(z) such that the likelihood of z is maximized under a chosen prior. Such techniques can perform exact log-likelihood computation and, therefore, input samples with low probability under the original training distribution can be flagged, and the output of the model for these samples can be regarded as unreliable.

Many of the current clinically integrated CVD related AI algorithms are semi-automated, relying partially on user input or editing. The approach described above, based on NF, may be employed to automatically detect erroneous user input, thus ensuring a robust performance of the algorithms when integrated in clinical workflows. For example, in a use case where the segmentation of the heart chambers is performed automatically but can be edited by the user, an NF based method may detect an erroneous / unlikely edit, thus, preventing a faulty prediction of a measure of interest (e.g., ejection fraction).

3.2 Non-invasive Assessment of Stable Coronary Artery Disease

The gold standard for quantifying the severity of CAD (Coronary Artery Disease) is the functional index Fractional Flow Reserve (FFR). FFR is measured invasively using a catheter but may be computed noninvasively using physics-based simulations performed on anatomical models reconstructed from medical images (computer tomography / X-ray coronary angiography). Physics-based simulations typically require long runtimes, which limits the clinical adoption. An alternative approach for computing FFR is based on machine learning models (Itu et al., 2016) (Figure 5). Therein, coronary anatomies are generated synthetically, and the ground truth values required for the training of the ML model are determined using the physics-based approach. FFR is predicted at each centerline point in the reconstructed anatomical model. ML-based FFR predictions were validated against the physics-based results and against catheter-based FFR in 125 lesions from 87 patients. An excellent correlation between the physics-based and machine-learning based predictions was obtained (0.9994, $p < 0.001$) and the Bland-Altman analysis found no systematic bias. The runtime, when compared to the physics-based approach was reduced by approximately 80 times, allowing for a real-time computation of FFR. Clinical studies subsequently successfully validated the ML-based FFR computation (Coenen et al., 2018).

The Instantaneous wave-Free Ratio (IFR) is another coronary diagnostic index employed for CAD assessment and decision making. A physics-based approach for computing IFR (c-IFR) was introduced by Passerini et al. (2017). The method relies on reduced-order fluid structure interaction hemodynamic computations and on coronary anatomical models reconstructed from X-ray coronary angiography. Two coronary angiograms at least 30° apart are selected and an end-diastolic frame is chosen for each of them. Next, the vessel centerline is manually traced on these frames by selecting a minimum of three points. The lumen

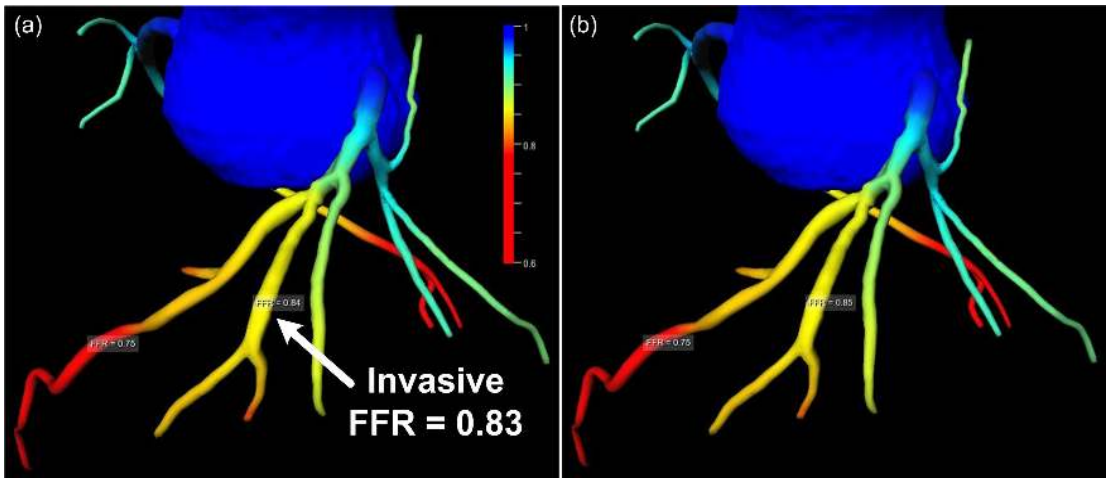

Figure 5. Example of a CT based coronary anatomical model reconstruction: (a) coronary tree color coded by physics based computed FFR values, with computed FFR = 0.84 and invasive FFR = 0.83, (b) coronary tree color coded by machine learning based computed FFR values, with computed FFR = 0.85

segmentation is then automatically computed and edited manually if required, and, next, the three-dimensional anatomical model is reconstructed from the lumen boundaries and the centerlines (Figure 6).

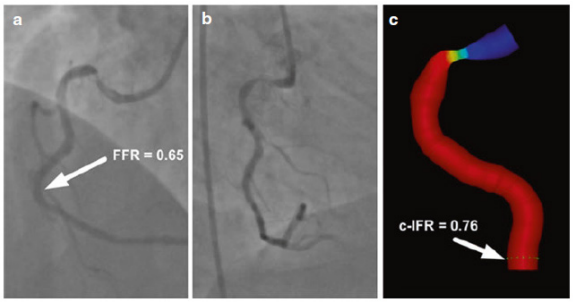

Figure 6. Example of an RCA lesion: (a) invasive FFR is 0.65 and c-IFR is 0.76, (b) end-diastolic frames of two different angiograms, (c) reconstructed coronary model with c-IFR color coded values (Passerini et al., 2017)

A hybrid decision making approach, which combines invasive FFR and c-IFR, was assessed against an invasive FFR-only approach in 64 patients and 125 lesions. Within the hybrid approach, lesions were deemed functionally significant for c-IFR < 0.86, non-significant for c-IFR > 0.93, and for intermediate values an invasive FFR based classification was performed. Overall, 43 lesions were functionally significant (FFR ≤ 0.8). When evaluated against the invasive FFR-only strategy, the hybrid approaches led to an accuracy of 96%, with the requirement of measuring FFR invasively in only 43 lesions (34%).

3.3 Treatment Planning in Aortic Coarctation

Computational modelling may also be used to predict the outcome of cardiovascular interventions, allowing for the evaluation of various treatment approaches and for the selection of the optimal treatment. A framework was proposed that combines CFD and ML based techniques for robustly and automatically personalizing aortic hemodynamic computations for the assessment of pre- and post-intervention aortic coarctation (CoA) patients from 3D rotational angiography (3DRA) data (Armstrong et al., 2019). The key features are: (i) a parameter estimation method for calibrating arterial wall properties, and inlet and outlet boundary conditions, to obtain computational results which match the patient-specific measurements, and (ii) a machine learning based pressure drop model which predicts pressure losses accurately for large variations of anatomical CoA models and flow conditions (Figure 7). The case series paper provided, besides a feasibility assessment, an initial validation of the framework against invasive measurements in three patients: a difference of less

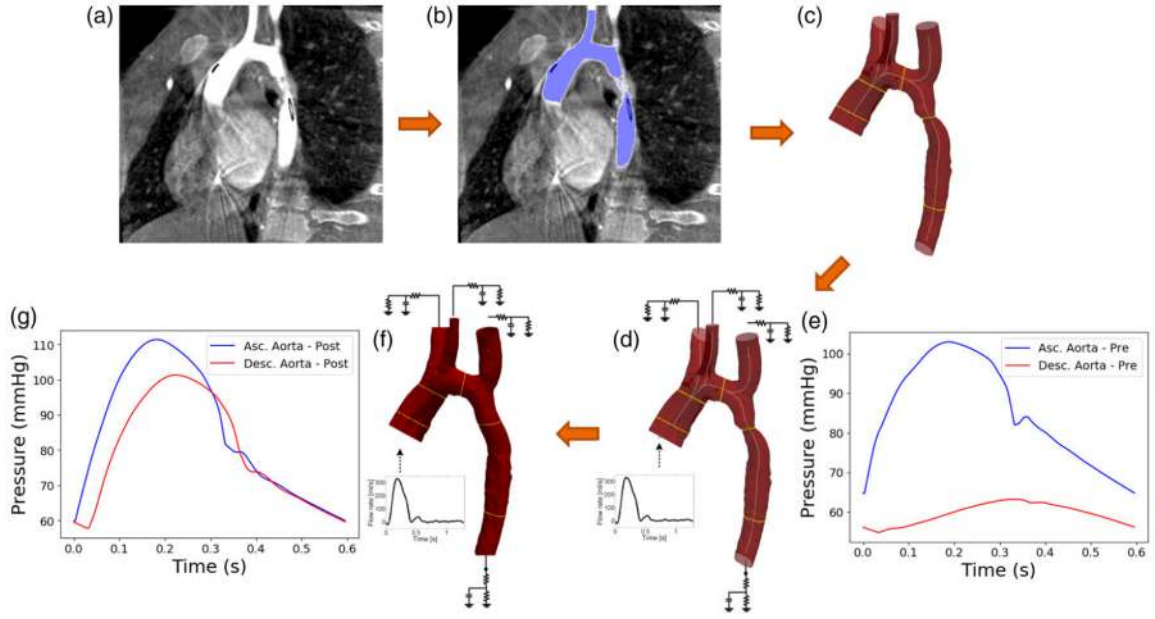

Figure 7. Physics based workflow for aortic coarctation assessment from 3DRA: (a) 3DRA (the arrow indicates the aortic coarctation), (b) aortic segmentation, (c) reconstructed 3D anatomical model, (d) definition of boundary conditions, (e) pre-stent hemodynamic results, (f) implantation of virtual stent, and (g) post-stent hemodynamic results (Armstrong et al., 2019)

than 5mmHg between computed and measured peak-to-peak trans-coarctation pressure drop was obtained for both the pre- and the virtual post-operative assessment.

4. Clinical Impact

AI based solutions can impact clinical care in various directions:

- doing what is right for the patient by focusing on the outcomes that matter for the patient: prioritize complex or acute cases (e.g., identify and prioritize acute coronary syndromes), avoid unnecessary interventions (e.g., minimize percutaneous coronary interventions while optimizing long term patient outcomes);

- transforming care delivery through improved efficiency and productivity: increase productivity through automation (e.g., CCTA based CAD gatekeeper: automatically detect patients without significant CAD), optimize clinical operations;

- expanding precision medicine through increased quality of care: patient and risk stratification (take the right clinical decision in terms of diagnosis and treatment), optimize the outcome (e.g., perform virtual treatment planning for aortic coarctation patient to select best treatment for minimizing the residual trans-coarctation pressure drop).

The recent theoretical and hardware related developments have lifted the potential of AI based solutions to a level where all of the above can be addressed. One of the most important remaining challenges is the access to data that cover the entire spectrum of possible pathological variations and combinations. The use of novel privacy preservation techniques may represent a very efficient and elegant answer enabling the access to more data. As discussed in section 2, the whole-body circulation model could thus perform robust predictions for patient stratification in dilated cardiomyopathy, left ventricular hypertrophy, mitral and aortic valve regurgitation and stenosis. While initial results of such solutions are promising, some challenging aspects still need to be further explored:

- maintaining a high level of accuracy when privacy related mechanisms are involved: currently a trade-off between accuracy and security is required;

- computational cost: although efficient in terms of security, current solutions suffer from increased run-time and computational cost.

Hence, it can be concluded that a trade-off exists between accuracy, security, and computational cost. While hardware developments will further reduce the computational cost, theoretical developments allowing for simultaneous high accuracy and security are required. A different approach for addressing the data challenge may be the use of synthetic data, which, in principle, does not raise any privacy concerns, but has other inherent challenges. Some examples are outlined in the previous sections: the AI models providing real-time assessment of the whole-body circulation and of coronary artery disease are trained on synthetically generated data. Furthermore, regardless of the amount of retrospectively collected data, to ensure that the models are trustworthy over time, they must continuously learn and readapt to changes that may occur in the statistical properties of the data caused by the shifts in patient population, treatment protocols or even symptoms.

The recent developments in AI explainability represent an important step forward for increasing the clinical acceptance of AI based solutions. To gain trust and ensure the acceptance by clinicians, the development of explainable AI models should actively involve end-users, i.e., clinical personnel, in the creation process. For the decision-making process related to aortic coarctation for example, being able to inspect the computed pressures at all locations along the aorta not only increases explainability, but also offers additional insight potentially leading to a better clinical decision.

While AI models are in general regarded as black-box models, clinical applications may be based on end-to-end or modular AI models. Modular applications divide the overall task of the application into smaller tasks, and inherently allow for a superior explainability, as the output of each task can be quantified, verified and validated. When referring to the diagnosis of coronary artery disease from computer tomography medical images, in an end-to-end approach, the AI model takes as input the 3D volume and outputs directly the diagnosis or the measures of interest (e.g., FFR). In a modular approach, different components are responsible for coronary artery centerline detection, lumen segmentation, and functional assessment.

Finally, the uncertainty of the model output is closely related to explainability and interpretability. AI model results should, whenever possible, be displayed together with uncertainty measures to increase the trust in the model, potentially leading to a higher acceptance and adoption of AI models in routine clinical practice. A hybrid decision making strategy could be devised, where AI based decisions with high confidence (low uncertainty) can be approved automatically without human interaction, whereas decisions with low confidence (high uncertainty) have to be reviewed or corrected by the physician. For example, referring again to the computation of FFR: if the confidence interval of the predictions does not contain the diagnostic threshold value of 0.8, the clinical decision may be taken automatically.

5. Conclusion

CVD is a major health problem around the world and will remain the most significant cause of mortality in the next two decades (Hu et al., 2016). So far, the use of AI has shown great potential in diagnosing, managing, and treating CVD. Cardiovascular imaging and its interpretation are likely to be changed by AI technologies, to enhance quality control, quantification, diagnosis, reporting, and workflow efficiency and ease of use. As the technology and its clinical validation progress, the integration of AI in clinical practice will further increase.

Herein, the present work has presented an overview of current approaches for privacy preserving AI, and has identified the potential next steps to be taken to reach maturity, i.e., clinical adoption: novel hardware solutions for reduced computational cost, and novel theoretical developments allowing for simultaneous high accuracy and security. Moreover, the knowledge gained from the use of AI models in routine clinical practice, will enable the development of more reliable and more complex and more accurate AI models. Regarding AI explainability and trustworthiness, two possible approaches have been detailed, i.e., the assessment of uncertainty and the integration with computational models (Niederer et al., 2019), which may reveal diagnostic information that otherwise would remain concealed.

Machine learning, deep learning, and AI in general are changing the way in which medicine is practiced. While remaining challenges are being addressed, physicians need to embrace and be prepared for the AI era, paving the way toward better diagnosis and precision medicine in cardiology and cardiovascular imaging.

Acknowledgments

The authors gratefully acknowledge the input of Costin Ciusdel, Alexandru Turcea, Diana Stoian, Alina Toma, Daniel Bunescu, Vasile-George Marica and Constantin Suciu in the preparation of the manuscript. The research reported in this paper was supported by a grant of the Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2019-1804, within PNCDI III.

REFERENCES

Abadi, M. & Andersen, D. (2016). Learning to Protect Communications with Adversarial Neural Cryptography, *ArXiv*, abs/1610.06918.

Armstrong, A., Zampi, J. D., Itu, L. M. & Benson, L. (2019). Use of 3D rotational angiography to perform computational fluid dynamics and virtual interventions in aortic coarctation, *Catheterization and Cardiovascular Interventions, 95*(2), 294-299.

Benjamins, J., Hendriks, T., Knuuti, J., Juarez-Orozco, L. & Harst, P. V. (2019). A primer in artificial intelligence in cardiovascular medicine, *Netherlands Heart Journal, 27*(9), 392-402.

Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A. & Torr, P. (2016). Fully-Convolutional Siamese Networks for Object Tracking, *ArXiv*, abs/1606.09549.

Bond, R., Novotny, T., Andrsova, I., Koc, L., Sisakova, M., Finlay, D., Guldenring, D., McLaughlin, J., Peace, A., McGilligan, V., Leslie, S., Wang, H. & Malik, M. (2018). Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms, *Journal of Electrocardiology, 51*(6S), S6-S11.

Bond, R., Rjoob, K., Finlay, D., McGilligan, V., Leslie, S. J., Knoery, C., Iftikhar, A., McShane, A., Tache, I., Biglarbeigi, P., Manktelow, M. & Peace, A. (2020). Near future artificial intelligence in interventional cardiology: new opportunities and challenges to improve the care of STEMI patients, *Journal of ESC Digital Health*. Corpus ID: 219659633.

Chabanne, H., Wargny, A. D., Milgram, J., Morel, C. & Prouff, E. (2017). Privacy-Preserving

Classification on Deep Neural Network, *IACR Cryptol. ePrint Arch*., 35.

Coenen, A., Kim, Y., Kruk, M., Tesche, C., Geer, J. D., Kurata, A., Lubbers, M., Daemen, J., Itu, L.M., Rapaka, S., Sharma, P., Schwemmer, C., Persson, A., Schoepf, U., Kępka, C., Yang, D. H. & Nieman, K. (2018). Diagnostic Accuracy of a Machine-Learning Approach to Coronary Computed Tomographic Angiography–Based Fractional Flow Reserve: Result from the MACHINE Consortium, *Circulation Cardiovascular Imaging, 11*(6), e007217.

Dey, D., Slomka, P., Leeson, P., Comaniciu, D., Shrestha, S., Sengupta, P. & Marwick, T. (2019). Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review, *Journal of the American College of Cardiology, 73*(11), 1317-1335.

Dowlin, N., Gilad-Bachrach, R., Laine, K., Lauter, K. E., Naehrig, M. & Wernsing, J. (2016). CryptoNets: applying neural networks to encrypted data with high throughput and accuracy. In *Proceedings of the 33rd International Conference on Machine Learning (ICML), 48* (pp. 201-210).

Duguay, T., Tesche, C., Vliegenthart, R. et al. (2017). Coronary Computed Tomographic Angiography-Derived Fractional Flow Reserve Based on Machine Learning for Risk Stratification of Non-Culprit Coronary Narrowings in Patients with Acute Coronary Syndrome, *The American Journal of Cardiology, 120*(8), 1260-1266.

Fischer, A., Klein, P., Radulescu, P., Gulsun, M., Ali, A., Schöbinger, M., Sahbaee, P., Sharma, P. & Schoepf, U. (2020). Deep Learning Based Automated Coronary Labeling for Structured Reporting of Coronary CT Angiography in Accordance with SCCT Guidelines, *Journal of Cardiovascular Computed Tomography, 14*(3), S21-S22.

Fratz, S., Chung, T., Greil, G., Samyn, M., Taylor, A., Buechel, E. V., Yoo, S. & Powell, A. (2013). Guidelines and protocols for cardiovascular magnetic resonance in children and adults with congenital heart disease: SCMR expert consensus group on congenital heart disease, *Journal of Cardiovascular Magnetic Resonance, 15*(1), 51.

Ghesu, F., Georgescu, B., Zheng, Y., Grbic, S., Maier, A., Hornegger, J. & Comaniciu, D. (2019). Multi-Scale Dep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(1), 176-189.

Gong, Y. & Poellabauer, C. (2018). Deep Obfuscation: Precise Masking of Sensitive Information to Protect Against Machine Learning Adversaries. In *Computer Security Applications Conference*.

Hall, J. (2011). *Guyton and Hall Textbook of Medical Physiology*, 12th ed. Saunders Elsevier.

Haq, I., Haq, I. & Xu, B. (2020). Artificial intelligence in personalized cardiovascular medicine and cardiovascular imaging, *Cardiovascular Diagnosis and Therapy*, DOI: 10.21037/cdt.2020.03.09

Hesamifard, E., Takabi, H. & Ghasemi, M. (2017). CryptoDL: Deep Neural Networks over Encrypted Data, *ArXiv*, abs/1711.05189.

Hu, J., Cui, X., Gong, Y., Xu, X., Gao, B., Wen, T., Lu, T., & Xu, F. (2016). Portable microfluidic and smartphone-based devices for monitoring of cardiovascular, *Biotechnology Advances*, *34*(3), 305-320.

Itu, L., Rapaka, S., Passerini, T., Georgescu, B., Schwemmer, C., Schoebinger, M., Flohr, T., Sharma, P. & Comaniciu, D. (2016). A machine-learning approach for computation of FFR from CCTA, *Journal of Applied Physiology, 121*(1), 42-52.

Kayvanpour, E., Mansi, T., Sedaghat-Hamedani, F. et al. (2015). Towards Personalized Cardiology: Multi-Scale Modeling of the Failing Heart, *PLoS ONE, 10*(7), e0134869.

Kipnis, A. & Hibshoosh, E. (2012). Efficient Methods for Practical Fully Homomorphic Symmetric-key Encrypton, Randomization and Verification, IACAR Cryptology ePrint Archive, Report 2012/637, 637.

Kobyzev, I., Prince, S. & Brubaker, M. (2020). Normalizing Flows: An Introduction and Review of Current Methods, *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning, *Nature, 521*(7553), 436-444. DOI: 10.1038/nature14539

Mansi, T., Passerini T. & Comaniciu D. (2019). *Artificial Intelligence for Computational Modeling of the Heart*. Academic Press.

Mathur, P., Srivastava, S., Xu, X. & Mehta, J. (2020). Artificial Intelligence, Machine Learning, and Cardiovascular Disease, *Clinical Medicine Insights. Cardiology, 14,* 1-9.

Naghavi, M., Abajobir, A. A., Abbafati, C. et al. (2017). Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016, *Lancet, 390*, 1151-1210.

Niederer, S. A., Lumens, J. & Trayanova, N. A. (2019). Computational models in cardiology, *Nature Reviews Cardiology, 16*(2), 100–111.

Obermeyer, Z. & Emanuel, E. (2016). Predicting the Future - Big Data, Machine Learning, and Clinical Medicine, *The New England Journal of Medicine, 375*(13), 1216-1219.

Orlandi, C., Piva, A. & Barni, M. (2007). Oblivious Neural Network Computing via Homomorphic Encryption, *EURASIP Journal on Information Security*, 1-11.

Passerini, T., Itu, L. & Sharma, P. (2017). Patient-Specific Modeling of the Coronary Circulation. In: Itu, L., Sharma, P. & Suciu C. (eds.), *Patient-specific Hemodynamic Computations: Application to Personalized Diagnosis of Cardiovascular Pathologies*, 61-88. Springer, Cham.

Pulerwitz, T. C., Khalique, O. K., Leb, J. et al. (2019). Optimizing Cardiac CT Protocols for Comprehensive Acquisition Prior to Percutaneous MV and TV Repair, *JACC: Cardiovascular Imaging*, *13*(3), 836-850.

Quionero-Candela, J., Sugiyama, M., Schwaighofer, A. & Lawrence, N. (eds.) (2009). *Dataset Shift in Machine Learning* (Neural Information Processing series). The MIT Press.

Richard, A., Mayag, B., Talbot, F., Tsoukiàs, A. & Meinard, Y. (2020). Transparency of Classification Systems for Clinical Decision Support, *Information Processing and Management of Uncertainty in Knowledge-Based Systems, 1239*, 99 - 113.

Romanelli, M., Chatzikokolakis, K. & Palamidessi, C. (2020). Optimal Obfuscation Mechanisms via Machine Learning. In *2020 IEEE 33rd Computer Security Foundations Symposium (CSF)*, (pp. 153-168).

Romiti, S., Vinciguerra, M., Saade, W., Cortajarena, I. A. & Greco, E. (2020). Artificial Intelligence (AI) and Cardiovascular Diseases: An Unexpected Alliance, *Cardiology Research and Practice, 5*, 1-8.

Rudin, C. (2018). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nature MIntelligence, 1*, 206-215.

Saltybaeva, N., Schmidt, B., Wimmer, A., Flohr, T. & Alkadhi, H. (2018). Precise and Automatic Patient Positioning in Computed Tomography: Avatar Modeling of the Patient Surface Using a 3-Dimensional Camera, *Investigative Radiology, 53*(11), 641–646.

Shokri, R. & Shmatikov, V. (2015). Privacy-preserving deep learning. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, (pp. 909-910).

Spevack, D., Karl, J.A., Yedlapati, N., Goldberg, Y. & Garcia, M. (2013). Echocardiographic left ventricular end-diastolic pressure volume loop estimate predicts survival in congestive heart failure, *Journal of Cardiac Failure, 19*(4), 251-259.

Thomas, H., Diamond, J., Vieco, A., Chaudhuri, S., Shinnar, E., Cromer, S., G., Narula, J., Johnson, C., Roth, G. A. & Moran, A. (2018). Global Atlas of Cardiovascular Disease 2000-2016: The Path to Prevention and Control, *Global Heart, 13*(3), 143-163.

Timmis, A., Townsend, N., Gale, C., Grobbee, R., Maniadakis, N., Flather, M., Wilkins, E., Wright, L., Vos, R., Bax, J. J., Blum, M., Pinto, F. & Vardas, P. (2018). European SoCiety of Cardiology: Cardiovascular Disease Statistics 2017, *European Heart Journal, 39*(7), 508-579.

Virani, S., Alonso, A., Benjamin, E. et al. (2020). Heart Disease and Stroke Statistics – 2020 Update: A Report from the American Heart AsSoCiation, *Circulation, 141*(9), e139-e596.

Vizitiu, A., Nita, C. I., Puiu, A., Suciu, C. & Itu, L. (2020). Applying Deep Neural Networks over Homomorphic Encrypted Medical Data, *Computational and Mathematical Methods in Medicine, 4*, 1-26.

Yan, Y., Zhang, J., Zang, G. & Pu, J. (2019). The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine?, *Journal of Geriatric Cardiology: JGC, 16*(8), 585-591.

Zhang, F., Kanik, J., Mansi, T., Voigt, I., Sharma, P., Ionasec, R., Subrahmanyan, L., Lin, B.A., Sugeng, L., Yuh, D., Comaniciu, D., & Duncan, J. (2017). Towards patient-specific modeling of mitral valve repair: 3D transesophageal derived parameter estimation, *Medical Image Analysis, 35*, 599-609.